

Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models

Martin Oswaldo Mendez*

Bioengineering Department,
Politecnico di Milano,
piazza Leonardo da Vinci 32, 20133 Milan, Italy
Fax: 390223993360
E-mail: martin.mendez@polimi.it
*Corresponding author

Matteo Matteucci

Electronics and Information Department,
Politecnico di Milano,
piazza Leonardo da Vinci 32, 20133 Milan, Italy
Fax: 390223993360
E-mail: matteucc@elet.polimi.it

Vincenza Castronovo and Luigi Ferini-Strambi

Sleep Disorder Center,
University Vite a Salute,
Ospedale San Raffaele,
Via Prinetti 29, I-20127 Milan, Italy
E-mail: castronovo.vincenza@hsr.it
E-mail: ferinistrambi.luigi@hsr.it

Sergio Cerutti and Anna Maria Bianchi

Bioengineering Department,
Politecnico di Milano,
piazza Leonardo da Vinci 32, 20133 Milan, Italy
Fax: 390223993360
E-mail: sergio.cerutti@polimi.it
E-mail: annamaria.bianchi@polimi.it

Abstract: An alternative DSS which models the behaviour of the Heart Rate Variability (HRV) signal linked to stable (NREM) and instable (REM) cerebral waves during sleep and a probabilistic model of the sleep stages transitions for decision was developed. Time-Varying Autoregressive Models (TVAMs) were used as feature extractor while Hidden Markov Models (HMM) was used as time series classifier. 24 full polysomnography recordings from healthy sleepers were used for the analysis and those were separated in two sets of

12 each: training and test set. The classification performance for the test set was specificity = 0.851, accuracy = 0.793 and sensitivity = 0.702.

Keywords: HRV; heart rate variability; sleep; DSS; decision-support systems; sleep staging; time-varying analysis; HMM; hidden Markov model.

Reference to this paper should be made as follows: Mendez, M.O., Matteucci, M., Castronovo, V., Ferini-Strambi, L., Cerutti, S. and Bianchi, A.M. (2010) 'Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models', *Int. J. Biomedical Engineering and Technology*, Vol. 3, Nos. 3/4, pp.246–263.

Biographical notes: Martin Oswaldo Mendez has an Engineering Degree in Electronics from the Tecnológico de Aguascalientes, Mex (2001), and has obtained an MSc in Bioengineering in 2003 with a Thesis on Non-Parametric and Non-Linear Analysis of the Heart Rate Variability during Apnoea Manoeuvres. In 2004, he obtained the PhD from the Department of Bioengineering of the Politecnico di Milano. Nowadays, he works on the Heart Rate Variability during Sleep and on related pathologies (such as Sleep Apnoea) by using parametric and non-parametric approaches as well as non-linear methods.

Matteo Matteucci's research is focused on learning machines (i.e., neural network, decision trees, mixture models, etc.). He has applied learning methods to different industrial and academic applications, becoming a reference source for this w.r.t. the local research community. Bayesian approaches to model adaptation and learning, neural models for biological signals interpretation (e.g., age prediction from heart rate variability, sleep staging, obstructive sleep apnoea recognition, lung cancer diagnosis), augmented and alternative language models, are just few examples of his activity in complex system modelling.

Vicenza Castronovo received a PhD Degree in Clinical Psychology and Psychotherapy from the University of Torino and a Master Degree in Psychology from the University of Parma. She practices as a Clinical Psychologist at Sleep Disorders Centre, Scientific Institute San Raffaele Milano. She is trained in behavioural sleep medicine and in particular in cognitive behavioural therapy for insomnia. Her activities also include research in sleep disorders in particular sleep apnea, insomnia, RLS and parasomnia. She is founding member and Honorary President of ESST (European Society of Sleep Technologists).

Luigi Ferini-Strambi received a Postgraduate Degree in Neurology from State University of Milan, Italy, in 1984 and an MD in 1980. At present, he is Director of the Sleep Disorders Centre, Department of Neurology, Scientific Institute H San Raffaele, Milan, Italy, and Associate Professor of General Psychology at University Vita-salute san Raffaele, Milan, Italy. He is former President of the Italian Association of Sleep Medicine. He is member of several scientific societies, including the Society of Autonomic Nervous System and the American Academy of Sleep Medicine. He is the Chair of the Membership Committee of the World Association of Sleep Medicine Society.

Sergio Cerutti is Professor in Biomedical Signal and Data Processing at the Department of Biomedical Engineering of the Politecnico di Milano, Italy. His research interests are mainly in the following topics: biomedical signal processing (ECG, blood pressure signal and respiration, cardiovascular variability signals, EEG and evoked potentials), cardiovascular modelling,

neurosciences, regulation and standards in medical equipments and devices. Since 1983 he has taught a course on Biomedical Signal Processing at Engineering Faculties. He is actually Fellow Member of IEEE and of EAMBES and Associate Editor of IEEE Trans BME. He is author of more than 400 international scientific contributions.

Anna Maria Bianchi received the Laurea Degree from the Politecnico di Milano in 1987. During the period 1987–2000, she was Research Assistant in the Laboratory of Biomedical Engineering of the IRCCS S.Raffaele Hospital in Milano, where her scientific and research activity was in connection with the Department of Biomedical Engineering of the Polytechnic University. Since 2001, she is Research Assistant in the Department of the Biomedical Engineering of the Politecnico di Milano. She is now Assistant Professor of Fundamentals of Electronic Bioengineering in the Biomedical Engineering School and of Biomedical signal and Data Processing in the PhD course.

1 Introduction

In past decades, the progress in mathematical modelling and the availability of computational resources paved the way for the automatic resolution of relevant and time-consuming tasks. These advances allow the analysis of large amounts of data and the development of Decision-Support Systems (DSS) to reduce repetitive and tiring works to human experts. In many different clinical disciplines, DSS may aid to establish feasibility of measurements and develop automatic screening procedures to analyse more data in a shorter time. A clear example of this, coming from sleep medicine is the usual practice to perform the sleep staging procedure through the visual scoring of polygraph traces that include many signals (i.e., electroencephalogram EEG, electromyogram EMG, electrooculogram EOG, respiration, and others) recorded along the whole night. Such a Polysomnography (PSG) evaluation is the gold standard procedure for sleep staging (Rechtschaffen and Kales, 1968) and, in general, for the diagnosis of main sleep disorders (Task Force, 1999). However, PSG presents some inconvenients such as the need for some dedicated equipment, dedicated sleep centres, specialised and expert technical personnel. Mainly, because of the reduced number of sleep centres, sleep diagnosis is a very expensive procedure and sometimes the waiting list for the patients is very long. As a consequence, many relevant sleep disorders remain underestimated and untreated generating high health deterioration.

All previous considerations lead to develop automatic procedures for the analysis of the sleep. Some efforts with fine results have been done to develop DSS for sleep screening based on the PSG or only on the EEG (Todorova et al., 1997; Agarwal and Gotman, 2001; Principe et al., 1989; Porée et al., 2006; Flexer et al., 2005). However, it has been observed that some peripheral measures, such as Heart Rate Variability (HRV), blood pressure and respiratory activity, present specific oscillatory patterns that are modulated in relation to the different sleep stages (Bonnet and Arand, 1997; Scholz et al., 1997; Penzel et al., 2003). For instance, during Non-Rapid Eye Movement (NREM) sleep, HRV presents spectral components well concentrated around the respiratory frequency; while during REM (rapid eye movement) sleep, HRV spectrum presents low-frequency components, which seem to be linked with the instability in the EEG

cerebral waves. Specific mathematical models could capture these typical variations and lead towards the development of a classifier to define a DSS for fast sleep screening.

HRV offers advantages such as high signal to noise ratio and ease of acquisition. These characteristics make HRV a very interesting signal for the development of DSS and it could be recorded at home as well as in usual medical environment. In addition, the spectral components of the HRV have a physiological meaning since these are closely related to the Autonomic Nervous System (ANS) (Task Force, 1996). Behaviour of the HRV has been largely explored during sleep, and many studies have been focused on its use for fast screening of sleep apnoea. The results obtained by those studies present high performance in the recognition of normal and pathologic sleep (Penzel et al., 2002). However, to extract the spectral features from the HRV, a method requires some specific features since HRV is non-stationary during sleep. Different approaches offer fine properties for non-stationary time series; some of them are Wavelets analysis, Empirical Mode Decomposition, Time-Frequency analysis and Time-Varying Autoregressive Model (TVAM) analysis (Cerutti et al., 2001).

The first step, for creating a good DSS, is to define the best approach able to extract representative features for the specific classification task from the raw data. In our case, we explored TVAMs since they are mathematical approaches that fit the data in a moving window updating the coefficients of the model for each new sample in the signal and using a forgetting factor for better adaptation. This characteristic gives high flexibility and adaptability to the TVAMs to describe different conditions of stationarity and non-stationarity that HRV signal can present during the different sleep stages (Bianchi et al., 1993). Then, a good estimation of the spectral components of the signal can be obtained from the coefficients and these can be used as features for classification as well.

Finally, the last step is the selection of the best approach that can help us to decide the sleep stage based on our observations or input signals. There are a large number of techniques such as Fuzzy Logic, Neural Networks and Discriminants, which could be used to classify objects or patterns based on the observation properties. We have selected Hidden Markov Models (HMMs) since they have interesting characteristics for problems that have an inherent temporality or a recognised pattern in time, such as REM–NREM sleep cycles. HMMs attempt to model a process where a sequence of events occurs and an intrinsic underlying pattern exists (Rabiner, 1989).

The aim of this study is to develop a DSS for sleep classification based on the analysis of HRV, which can be recorded more easily than EEG. In particular, we concentrate on an REM and NREM classification good enough for fast sleep screening, which could be easily carried on in environments out of sleep centres. TVAMs have been used as feature extractor while HMM was used as time series classifier.

2 Material and methods

2.1 Clinical protocol

This study was carried out in collaboration with the San Raffaele Hospital in Milan, Italy. Only healthy subjects with normal sleep efficiency were selected for this study. Clinical aspects, incidence sleep aspects and polysomnography records have been evaluated by a physician expert in sleep medicine. Age of the subjects ranges between 40 and 50 years.

Subjects had a body mass index less than 29 kg/m^2 and the database for this study includes 24 subjects. 19 males and 5 females formed the database. The range of apnea-hypopnea index (AHI) was 0 for all subjects, all of them were drug-free and each subject participated with one night recording. Mean sleep efficiency was 83%.

All experiments were conducted in a special department at the sleep clinic of the San Raffaele Hospital. The sleep centre has four bedrooms. The temperature is regulated by mean of an air-conditioner and bedrooms are completely dark and windowless.

Sleep evaluation was assessed by polysomnography in the sleep centre. This consisted of two central (C3, C4) and two occipital (O1, O2) EEG derivations, electrooculogram (EOG, left and right), electromyogram (EMG, submental), leg movement (tibialis anterior muscles left), airflow, thoracic and abdominal efforts, oxygen saturation and Electrocardiogram (ECG). These signals were collected according to the standardised criteria (Rechtschaffen and Kales, 1968). The acquisition system was a Heritage Digital PSG Grass Telefactor and all data were acquired with 128 Hz as sampling rate.

Polysomnographic data were scored in epochs of 30 s each, according to the gold standard criteria established by Rechtschaffen and Kales (1968). As a result of this procedure, the hypnogram, which reports the different sleep stages across the sleep time, was obtained. For the study, only the hypnogram and the ECG signal were used.

The database was divided in two groups, each containing 12 different subjects: training group and test group. Each group contains approximately 10,000 epochs. The training set was used to build up and test our algorithm of classification (HMM). The test group was used to measure the performance of the algorithm. Clinical hypnogram was used to quantify the algorithm performance.

2.2 *HRV extraction*

The ECG was extracted from the polysomnography data; *R* peaks were detected from ECG using a derivative built and tested algorithm, and parabolic interpolation was added to overcome the limitation owing to a low sampling rate. Distances between consecutive *R* peaks were evaluated. This procedure gave as result the tachogram. Some *R* peaks were misdetections and some ectopic beats were found in the ECG. Then, ECG and tachogram were plotted together to observe clearly the erroneous detections. Whereas a beat or a series of beats were misdetections, these were manually corrected and the new tachogram recalculated.

2.3 *Arousal searcher*

An automatic searcher for autonomic arousal was developed. This algorithm searches the arousal events from the HRV (Sforza et al., 2000). To define an arousal, a sequence of consecutive beats must obey the next criteria; at certain beat occurring at time t :

- Beat at time $t + 2$ must be at least 5% shorter than the beat at time t
- Beat at time $t + 4$ must be at least 10% shorter than the beat at time t
- Beat at time $t + 6$ must be at least 15% shorter than the beat at time t

- At least one beat between time $t + 6$ and $t + 15$ must be at least 3% longer than the beat at time t
- If all previous criteria are true, there is an arousal at time t and the new arousal search starts again 20 beats later, this procedure is repeated for the total length of the time series.

2.4 Feature extraction

Extracted and corrected time series presented both steady-state periods and segments with rapid changes. Those fast variations produce non-stationarities in the signal, and to overcome this problem a TVAM was used since being suitable to analyse non-stationary signals. In TVAMs, an adaptive model self-adjusts its transfer function in relation to a prediction error. The prediction of the current output for an *all pole system* may be obtained as follows:

$$\hat{y} = -\sum_{k=1}^P a_k y(n-k),$$

where n denotes the time, a_k is the coefficient of the model, P is the filter order (in the present application P was chosen to be 8) and $\hat{y}(n)$ denotes the prediction output. The prediction error is evaluated as:

$$e(n) = y(n) - \hat{y}(n) = y(n) + \sum_{k=1}^P a_k y(n-k).$$

Self-adjustment uses a Cost Function (CF), which determines the algorithm performance. The goal in adaptive models is to minimise this CF (i.e., in the sense of least square) at each time n :

$$CF_p(n) = \sum_{k=1}^n \lambda^{n-k} |e(n)|^2$$

where k represents the observation interval, λ (it was selected 0.98 for this application) is the forgetting factor and $1/(1 - \lambda)$ represents the memory of the algorithm. λ weights the vector of prediction error giving more importance to recent errors, λ^{n-k} is always less than one. Self-adjustment and forgetting factor are desirable when we work with non-stationary sequences; when we deal with non-stationary signals, Recursive Least Squares (RLS) is the most adequate algorithm to update the filter parameters (Bianchi et al., 1993).

Thereafter, the time evolution of the HRV spectrum was directly computed from the obtained TVAM parameters. A whole night, time-variant model was calculated for each subject on a beat-by-beat basis using eight coefficients as model order. Model order selection criterion was the minimum number of coefficients that could fit the HRV signal to simplify the analysis and to assure only one pole in each spectral band of clinical interest (defined in the following according to the physiological interpretation reported in Task Force (1996)). RLS algorithm was used to update autoregressive parameters. The forgetting factor 0.98 (window with 50 beats) for all the subject's records was chosen. Forgetting factor represents the model memory (Bianchi et al., 1993).

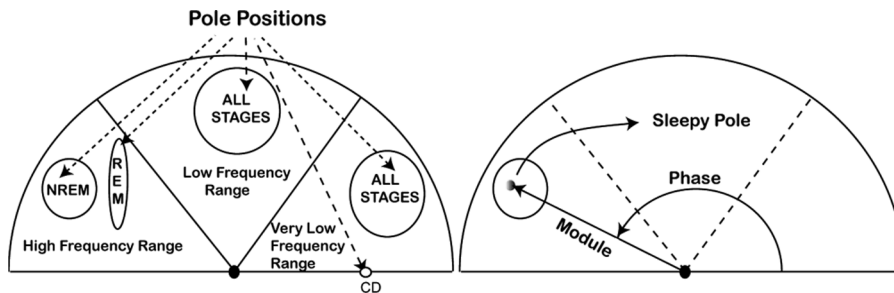
Power spectrum was obtained beat-by-beat and computed from the estimated time-varying autoregressive parameters for each time series. Spectral components may be obtained by calculating the power spectral contribution of each pole of the model in each band or only by subdividing the total spectrum in the specific bands. The second option is computed faster and it is expected that no band overlapping will take place. Therefore, from each estimated spectrum, we computed the following time-variant spectral indexes of the HRV:

- TP; Total Power (0.003–0.6 Hz)
- VLF; very low frequency component (0.003–0.02 Hz)
- LF; low-frequency component (0.02–0.15 Hz)
- HF; high-frequency component (0.15–0.6 Hz)
- LF/HF; Low to high frequency components ratio.

In addition to the classical spectral indexes, two new indexes were extracted from the model parameters. The modulus and the phase of the pole in the high-frequency band were extracted; this was possible since only one model pole might fall in each band due to the main HRV oscillations: offset, VLF, LF and HF. It was expected that the beat-by-beat space evolution of the high-frequency pole was characteristic of REM and NREM sleep stages. In the following, this new parameter will be named as *Sleepy Pole* and is defined by (see Figure 1):

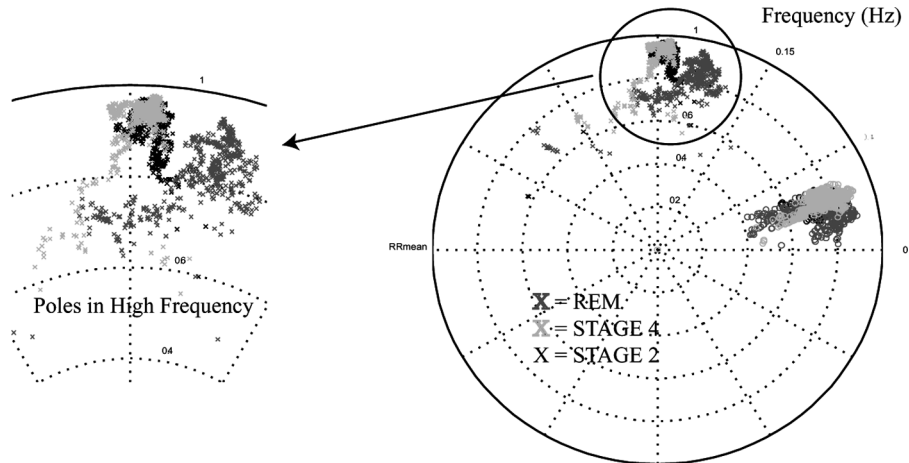
- the modulus of sleepy pole
- the phase of sleepy pole.

Figure 1 Scheme of pole positions during the different sleep stages. Definition of the sleepy pole indexes: modulus and phase



In case of two poles were present in the same band (i.e., HF band), the pole with the higher modulus is selected as the representative pole. From the *R*-peak time series, an average value for HRV was obtained each 30 s. The same procedure was carried out for all spectral indexes of the HRV and phase and modulus of the sleepy pole as well as for their respective differentials. From mean time series, wake epochs were removed as well as the epochs containing an autonomic arousal before REM and NREM classification and those epochs were again added after class separation procedure in the respective time location. This procedure is because the study only focuses on following the REM–NREM dynamic during the sleep time. Figure 2 shows an example of the sleepy pole behaviour during sleep Stages 2, 4 and REM.

Figure 2 Time-varying spectral analysis of RR series in REM, Stages 2 and 4 during sleep. Representative pole position for each sleep stage in the z -plane are REM = dark grey, Stage 2 = clear grey and Stage 4 = black



2.5 Selection and transformation of the features

Feature selection is one of the most important steps in pattern recognition. This procedure involves different tasks that simplify class separation for the decision system and prevent the *curse of dimensionality* in estimating the *a posteriori* distribution for the classification performed. If we select the best features that separate the classes and eliminate features with low informative content, then the algorithm efficiency increases and a better classification is obtained.

Different procedures can be used to select the best set of features for the specific application. Notice that the features are selected with specific application for the problem. Any feature has different information for a specific classification task and there are some different ways for selecting the best set of features. The first decision is to choose the features by simple observation. Selection can also be evaluated by statistical analysis of the features, based on either statistical differences or we could use WRAP methods. WRAP methods consist in selecting the features based on the classifier performance for each group of those (Kohavi and John, 1997).

In this work, the test of Wilcoxon/Mann-Whitney was applied to the features to find those features that were statistically different between NREM and REM sleep. Before applying Wilcoxon/Mann-Whitney to select the features, those were normalised to eliminate the subject inter-variability. Normalisation transforms a series of number to attain specific statistical properties such as limit values, variance, or average. Therefore, for each time series, the asymmetry and the outlier presence were evaluated through Fischer coefficient and box plot, respectively. Selection of the transformation was decided on the basis of asymmetry and outlier characteristics. After normalisation and transformation, Wilcoxon/Mann-Whitney was applied to the new time series. Only the features derived from HR power did not show statistical differences.

Finally, each feature of each subject was quantised to generate the alphabet of symbols consisting of numbers from 0 to 100 that were obtained by dividing the amplitude of the feature range in 100 equispaced intervals.

2.6 Hidden Markov Model

Hidden Markov Models (HMMs) have great use in problems that have an inherent temporality or a recognised pattern in time such as the REM–NREM cycle. HMMs attempt to model a process where a sequence of events occurs and an intrinsic pattern exists. A Markov process is a process that moves from state to state (i.e., Wake, NREM and REM sleep) depending only on the previous n states, here a state could be a sleep stage. This process is called an *order n* Markov model, where n is the number of previous states affecting the choice of the present state. The simplest Markov process is a first-order process, where the choice of the next state is made only on the basis of the previous one. Let us consider a sequence of states at successive time steps:

$$\mathbf{w}^T = \{w_1, w_2, \dots, w_T\}$$

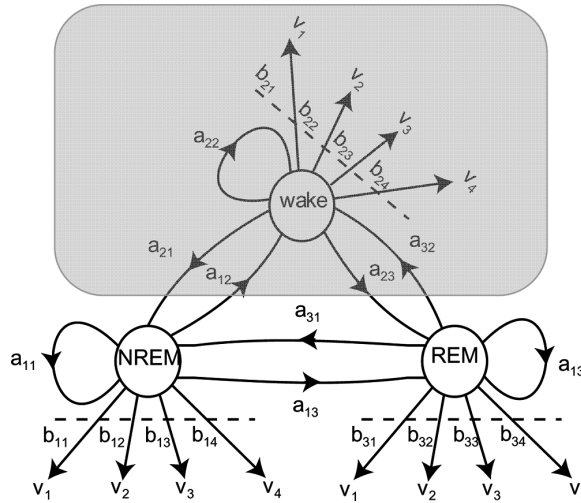
where T denotes the sequence length and w_i is the i state. For example, we could have a sequence of states described as $\mathbf{w}^6 = \{w_1, w_3, w_2, w_2, w_1, w_3\}$, where we notice that the system can remain in the same state at different steps, and not every state needs to be visited.

In Markov models, the production of any sequence is described by *transition probabilities* $P(w_j(t+1) | w_i(t)) = a_{ij}$. This means that a_{ij} is the time-independent probability of having state w_j at step time $t+1$ given the state w_i at time t . Figure 3 shows all possible first-order transitions among three states. Note that we hypothesise three states; however, wake is not included in this work and the actually used HMM is a two-state model.

Notice that for a first-order process with M states, there could be M^2 transitions among states, because it is possible, in the general case, for any state to follow any other. Furthermore, notice that these probabilities do not vary in time; this is an important (although often unrealistic) assumption. If a first-order discrete time Markov model, at any step t is in a particular state w_i , the state at step $t+1$ is a random function that depends only on the state at step t and the transition probabilities. The M^2 transition probabilities may be collected together in a matrix, which is named *State Transition Matrix*.

In practice, however, we do not have access to direct sleep stages, but we have all the parameters that we can extract from the physiological systems. In our particular case, spectral parameters extracted from the HRV signal will be the only observable piece of information for us and we are interested in an algorithm to classify the sleep stages from the HRV features and the Markov assumption. We assume that at every time t during sleep the system is in some state w_i and we also assume that it emits some visible symbols such as the sympathovagal balance level. Then, the observed sequences of symbols are probabilistically related to the hidden process states. One can model such a process using a HMM, where there is a Markov model not visible and a set of observable symbols that are related somehow to the hidden states.

Figure 3 General scheme of a Hidden Markov Model with three discrete states wake, NREM and REM, w_i , and their transition probabilities, a_{ij} . The emitted symbols (value of a feature) at each state are represented by v_k , and their respective probability of appearing at each discrete state are given by b_{ik} . The dashes lines represent the separation between the observable and hidden part of the model. The shadow area was not used in the present study



Let us define a particular sequence of visible symbols as $\mathbf{v}^T = \{v_1, v_2, \dots, v_T\}$ and thus we might have $\mathbf{v}^6 = \{v_4, v_1, v_1, v_5, v_2, v_3\}$. Then, we can describe our model as follows: in any state w we have a probability of emitting a particular visible state v_k . We denote this probability $P(v_k | w_j) = b_{jk}$. Because we have access only to the visible symbols and not to the w_i states, such a model is called HMM (see Figure 3) (Duda, 2001).

The connection between the hidden states and the observable symbols represents the probability of generating a particular observed symbol, given that the Markov process is in a particular state. This requires another matrix termed *Emission Matrix*. This matrix contains the probability distribution of the observable symbols given a particular hidden state. Since we are using probabilities, the sum of the probabilities for all symbols in a state must be 1 and this turns out in:

$$\sum_j a_{ij} = 1 \quad \forall i$$

and

$$\sum_k b_{jk} = 1 \quad \forall j.$$

HMMs can be used to solve three kinds of problems:

- *The evaluation problem.* It supposes that we have an entire HMM, i.e., State Transition Matrix (a_{ij}), Emission Matrix (b_{jk}) and the vector of initialisation π . Then, we want to know the probability that a particular sequence of observed states \mathbf{v}^T was generated by the model.

- *The decoding problem.* In a few words, it is to find the most probable sequence of hidden states \mathbf{w}^T given some observations \mathbf{v}^T .
- *The learning problem.* This problem deals with calculating the Markov model (State Transition Matrix and Emission Matrix) that have generated a sequence of observations.

Each probability in the state matrix and in the emission matrix is assumed to be time-independent. It is to say that the matrices do not change in time as the system evolves. In practice, this is one of the most unrealistic assumptions of the Markov models about real processes. For a deeper analysis of HMM and related problems, please refer Rabiner (1989).

2.7 *Estimation of the classification performance*

The learning procedure was carried out using only the training set (12 whole night recordings). From this set, the coefficients of the emission matrix and transition matrix for each feature were evaluated. It is to say that we generated one HMM model for each feature. The mean classifier performance for each feature was evaluated by Leave-One-Out (LOO) cross-validation technique applying the decoding algorithm (the best path of states that describes the observed physiological parameters) (Duda, 2001). This means that the transition and emission matrices were obtained from the training set for each feature by leaving one recording out and calculating the statistical measures of classification accuracy, sensitivity and specificity. Then, again from the whole training set another recording (never the same) was left out and the transition and emission matrices and measures of classification were again evaluated. This procedure was repeated until each recording was left out once. Finally, we obtain the mean performance for the 12 trials. For REM–NREM sleep classification, only those features with the best accuracy were used:

- Mean RR
- $\text{Log}(\text{Sqrt}(\text{VLF}))$
- $\text{Sqrt}(\text{Pole in HF (modulus)})$
- Pole in HF (phase).

The final state (REM or NREM) was decided by taking the state with major occurrence from the former features at each epoch. Finally, we classified all recordings of the test-set (12 recordings never seen for the algorithm). In this case, we used only the decoding algorithm (using the Transition and Emission matrices from the training set) for finding the best path of states that generated the sequence of observation in each subject.

3 **Results**

Classification of the NREM–REM sleep was carried out by applying a TVAM as feature extractor and a HMM as classifier. Twelve recordings were used to develop and train the classifier and other 12 to test it.

From the top to the bottom, Figure 4 shows RR mean, VLF, sleepy pole modulus, sleepy pole phase, sleep profile, NREM–REM obtained with HMM and the gold standard hypnogram for one subject. Medical hypnogram was simplified in wake, NREM (Stages 1–4) and REM sleep. In the sleep profile and hypnogram: 0 corresponds to wake, 3 is NREM and 5 REM. In addition, for the sleep profile, 6 represents the autonomic arousals found in the HRV. One can observe a high correspondence between the gold standard and the classification calculated by the algorithm. RR intervals show high variability during REM sleep as well as the phase of the pole. VLF tends to be higher during REM while the pole modulus decreases. Note that wake epochs were excluded from classification a priori and replaced afterwards.

Figure 4 Sleep profile of a normal subject obtained from HRV. From the top to the bottom, RR intervals, Very Low Frequency (VLF), modulus of the pole in high frequency, phase of the pole in high frequency, sleep profile evaluated by Hidden Markov Model and hypnogram given by the physicians. In sleep profile and hypnogram 0 means wake, 3 means NREM, 5 is REM and 6 are the arousal obtained by the automatic algorithm. Wake stages were excluded for the classification a priori and reallocated afterwards (see online version for colours)

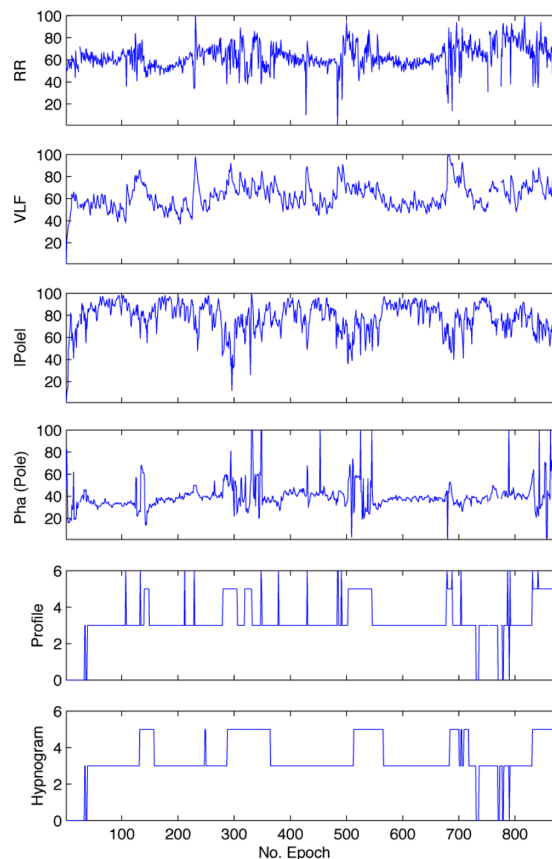


Table 1 shows the performing classification on training set, by LOO, and on test set. The algorithm presents fine performances in the classification of NREM and REM sleep from HRV parameters. Accuracy is around 80% in both training and test sets.

Table 1 Decision performance on the training set and test set by HMM classifier

<i>Set (%)</i>	<i>Specificity</i>	<i>Accuracy</i>	<i>Sensitivity</i>
Training	89.9	76.4	69.2
Test	85.1	79.3	70.2

4 Discussion

An alternative approach for identifying NREM and REM sleep from the HRV is presented. This approach pretends to be the first step of a complete sleep screening. The DSS is based on the patterns that characterise the ANS function during sleep, giving a continuous description of the human sleep. The approach presents several distinct characteristics. The most important one is the evaluation of a sleep profile based only on the projections of the central controls towards peripheral systems. This allows the use of signals of easy acquisition such as the surface electrical activity of the heart. We used algorithms that seem to be natural for evaluating sleep based on the time evolution of the HRV. TVAMs are used to generate a model in a beat-by-beat basis. This model presents specific characteristics for REM and NREM sleep. Then, HMM evaluates the most probable state sequence based on those characteristics and the probability of passing from one state to another in the successive step. Notice that both algorithms generate time-dependent models based on the own time series characteristics.

Some studies have developed automatic classifiers obtaining similar fine results. However, most of the efforts made until now have been based on the EEG (Todorova et al., 1997; Agarwal and Gotman, 2001; Principe et al., 1989; Porée et al., 2006; Flexer et al., 2005). When we think of alternative support systems for fast screening out the sleep centre, it is important to identify signals that satisfy some characteristics of robustness and easy acquisition. HRV presents such ease of acquisition that gives more comfort to the user as well as high signal to noise ratio that results in a low sensibility to the noise.

TVAMs present some interesting characteristics. One of them is the ability for adjusting the model parameters at each new beat along time. We can visualise this approach as the best linear polynomial that fits a data time window that runs across the data. From the polynomial parameters, it is possible to obtain the spectral components at each sample data, and consequently the extraction of different features that have physiological significance as well as valuable information to identify specific events. In addition, this approach does not need an equispaced sampling frequency as those approaches based on the Fourier transform. TVAMs have an elevated computational efficiency and a reduced complexity; these characteristics become fundamental when we think about implementing them in embedded systems or directly in hardware.

We may relate physiologic mechanisms with the TVAMs in four basic steps:

- cardiac control mechanisms (regulated by the ANS) change on a beat-by-beat basis according to physiological levels
- those variations are reflected as fluctuations on the generated beat series

- time-varying models are able to follow sample by sample the dynamic of a time series
- time-varying models could be able to follow beat-by-beat the dynamic of the ANS in the most diverse conditions.

The application of time-varying models to the HRV signal seems thus natural for monitoring physiological cardiac control mechanisms and for the diagnosis of health status. However, as any system in the real world, some difficulties exist that must be taken into account. HRV or, better for us, ANS reacts depending on the body necessities. Sympathetic tone increases in conditions of psychological or physiological stress to deliver enough oxygen and nutrients to the organs (Rhoades et al., 2006). The velocity of reaction depends directly on stress, physical condition and physiological limits. When these changes require a fast reaction, we found transitories on HRV produced mainly by sympathetic activity. On the other side, parasympathetic flow produces antagonistic changes. Moreover, not necessarily a high participation of one produces a low activation of the other one. It is surprising how this control is able to maintain in the best condition the body homeostasis. For example, in a simple rest-stand or rest-tilt test, very rapid changes on HRV can be observed (Furlan et al., 2000). Further, during an arousal from sleep, rapid changes in the HRV are caused by a withdrawal on parasympathetic tone and an increase on sympathetic activity (Sforza et al., 2000; Somers et al., 1993). Also, during normal sleep, the changes from deep sleep to states with higher cerebral activity such as Stage 2 or REM, rapid changes can be observed on HRV. Not only during normal conditions rapid changes are generated on HRV, we can also find sudden reaction of the ANS during pathologies such as obstructive sleep apnoea (Clary et al., 1995), central apnoea (Somers et al., 1993), and periodic leg movements (Sforza et al., 2005).

Besides autoregressive models seem appropriate tools to extract the spectral features of the HRV, they present some difficulties that it is necessary to take into account to obtain their best performance. TVAMs adjust the polynomial model coefficients in a minimum square mean sense based on the prediction error. This procedure assures that we obtain the best model for each time series for a given order, but to obtain the best performance of a time-variant model, it is necessary to select correctly two parameters: model order and forgetting factor. Different criteria have been developed to decide the best model order. Akaike criterion is one of the most widely used for this purpose. This option is suitable when we work with short time series or with some grade of complexity; for example, if we want to analyse HRV during sleep Stage 2, it is suitable to think in a possible mean model, since the statistical properties, in the different REM–NREM cycles, are more or less the same. However, if a transitory takes place, a different model order would be necessary. Then, the selection of the model order based on Akaike criterion turns out to be not suitable for sleep analysis if whole night series must be analysed on a beat-by-beat basis.

Forgetting factor is the second important parameter that it is necessary to select, this parameter is related to the convergence time of the filter and the importance of the past data. If we select a low forgetting factor, we consider only recent values, on the contrary, if a high forgetting factor is selected the contribution of the new data is masked by the past data, and as a consequence, the filter adaptation is very slow.

At this point, some questions could help us to solve the selection of both the best model order and forgetting factor. We are interested in finding the values for which we

can model or fit the sleep time series in all different circumstances and within the most variable statistical properties.

- What are the minimum and maximum number of coefficients necessary to describe the HRV signal in the most variable conditions?
- How important is the new beat to describe sleep stages?

It is important to take a look at the past; Authors in Scholz et al. (1997), Bianchi et al. (1993), Cerutti et al. (2001), Mainardi et al. (1997), Baselli et al. (1987) and Blasi et al. (2003) have applied model orders between 8 and 20, and forgetting factors between 0.96 and 0.995. This offers a reduced range to select the parameters. As first an intuition, selection of a higher model order could be a good option, because we can fit the HRV data in any condition, even if it is very complex. However, a problem arises when the signal is very regular or extremely stationary and over-fitting occurs. This situation is not desirable, since no realistic estimation of the data is obtained. On the other hand, when the lower order is selected, there is a risk of not modelling the dynamics of the signal. How one can appreciate, it is difficult to select the parameters to solve this paradigm, we adopted the following criterion.

- a normal person spends about 50% of sleep time in Stage 2
- adaptation speed based on a dynamic of at least 50 samples (beats) seems optimal, this corresponds to a forgetting factor of 0.98.

We tested the algorithm during all sleep stages, without taking care about the possible transitories (caused by stage change, arousal, movements, and so on) changing model order between 8 and 12 coefficients and forgetting factor between 0.975 and 0.992. We observed that during steady conditions, which occur during NREM sleep, very often we had over-fitting with model order equal to 12. Forgetting factor lesser than 0.98 gave too so much importance to the new entrance beat and the information contained in the past data is lost. Forgetting factors higher than 0.992 do not capture new information. In addition, we thought that if we left one couple of poles for each characteristic oscillation that appears in the HRV signal, the model might generate one pole pair for each spectral component that characterises our physiological system. In other words, two poles for the continuous component, two for VLF, two for LF and two for HF. Then, minimal order filter was selected (8 coefficients) together with forgetting factor equal to 0.98.

From the classifier perspective, HMM is an interesting mathematic approach to generate descriptive models of stochastic signals, such as the HRV during sleep. Different to other approaches that could be useful for the sleep stage decision such as discriminants, support vector machines and neural networks, HMM carries temporal information about the intrinsic dynamic of the process under analysis. However, in normal sleep, one HMM for each REM–NREM cycle is the ideal case to define more accurately the sleep stages since each REM–NREM cycle presents different structure, which means that NREM time is larger than REM one at the first REM–NREM cycle, while NREM time is smaller than REM one at the last REM–NREM cycle. But, for applicative situations, one mean HMM could be enough for classifying REM–NREM.

Only a few studies appear in the literature where sleep staging is evaluated from HRV. Lewicke et al. (2008) presents a study in which support vector machines,

neural networks and learning vector quantisation are used to separate wake and sleep in children. The results have a good agreement level with the gold standard (Rechtschaffen and Kales, 1968). Another interesting study was presented by Redmond et al. (2007) in which wake, NREM and REM sleep separation is evaluated. Authors focus on the application of simple procedures for both feature extraction (Fourier analysis) and stage classification (Discriminants). A good agreement level was obtained in this study, especially between sleep and wake. Finally, Watanabe and Watanabe (2004) presented a study that is very similar to the current one. Also, the results were good for wake and NREM separation, however REM classification did not present good results.

With respect to those previous studies, one criticism to our study is the exclusion of the wake stage. However, this study presents a first step towards a three-state decision system, wake, NREM and REM. When we want to separate wake state from our DSS, the performance for REM–Wake separation is relatively low, since HRV presents similar behaviour during REM sleep and quite wake state, also noted in the Redmond study (Redmond et al., 2007). This situation also produces confusion when the scoring is only evaluated from EEG (Flexer et al., 2001). In clinics, the EMG and EOG separate REM and wake easily since muscular atonia and eyes movements are typical characteristics of REM sleep. However, Flexer et al. (2001) discussed that even EMG presents serious problems to separate REM and wake.

One possible solution to improve DSS performance is to generate multivariate features for HMM to find out a possible relation that helps us to separate Wake–REM. In addition, we could integrate HRV with simple accelerometers or even better with EMG sensors that could support in the wake-state decision. Another possible solution could be a waterfall procedure, Separation of Wake–REM from NREM (high-low cerebral activity separation with REM and wake representing high brain activity and NREM low brain activity). Thereafter, Wake–REM separation could be carried out easily. Note that it is very interesting how the HRV behaviour is directly related to the brain activation. Unfortunately, we were not able to carry out this experiment since only a few wake epochs were present in our database and some recordings do not present or present a few wake epochs before and after sleep time.

5 Conclusions

Peripheral signals carry important information about sleep behaviour. The influence that central nervous system has over all body are useful to evaluate some intrinsic mechanisms such as the circadian rhythms and sleep stages. Application of simple and manageable mathematic approaches, such as TVAMs and HMMs to those peripheral signals, approximate finely a sleep profile that is very similar to the standard one. The results showed a good agreement with the gold standard staging. It is important to note that this DSS does not try to reproduce the gold standard criteria for sleep diagnosis, but tries to evaluate sleep quality on the basis of the harmony of REM–NREM evolution. The DSS developed takes advantage of the high relationship between stable and instable sleep and its projections on the peripheral systems. This DSS could help as support decision tool for expert scorers when the visual classification of a sleep stage is difficult owing to the complex interpretation of the polysomnographic traces. The use of a single signal derived from ECG with favourable characteristics in terms of recording and noise sensitivity, rather than a set of signals, as required for the clinical polysomnography,

are of some interest to the DSS presented here. Among the most important characteristics, we can find: real-time evaluation, easy software implementation and low computational costs. Therefore, the DSS could be part of wearable devices or become an independent system. Furthermore, this DSS could be easily coupled with the HRV obtained from the photoplethysmography, since the DSS measures the variations in heart rate and it does not depend on other characteristics.

Acknowledgement

This paper was partially supported by the HeartCycle Project ICT FP7 216695 of the European Community.

References

- Agarwal, R. and Gotman, J. (2001) 'Computer-assisted sleep staging', *IEEE Trans. Biomed. Eng.*, Vol. 48, No. 12, pp.1412–1423.
- Baselli, G., Cerutti, S., Civardi, S., Lombardi, F., Malliani, A., Merri, M., Pagani, M. and Rizzo, G. (1987) 'Heart rate variability signal processing: a quantitative approach as an aid to diagnosis in cardiovascular pathologies', *Int. J. Biomed. Comput.*, Vol. 20, pp.51–70.
- Bianchi, A.M., Petrucci, E., Signorini, M.G., Mainardi, L.T. and Cerutti, S. (1993) 'Time-variant power spectrum analysis for the detection of transient episodes in HRV signal', *IEEE Trans. Biomed. Eng.*, Vol. 40, No. 2, pp.136–144.
- Blasi, A., Jo, J., Valladares, E., Morgan, B.J., Skatrud, J.B. and Khoo, M.C.K. (2003) 'Cardiovascular variability after arousal from sleep: time-varying spectral analysis', *J. Appl. Physiol.*, Vol. 95, No. 4, pp.1394–1404.
- Bonnet, M.H. and Arand, D.L. (1997) 'Heart rate variability: sleep stage, time of night, and arousal influences', *Electroencephalogr Clin. Neurophysiol.*, Vol. 102, No. 5, pp.390–396.
- Cerutti, S., Bianchi, A.M. and Mainardi, L.T. (2001) 'Advanced spectral methods for detecting dynamic behaviour', *Auton. Neurosci.*, Vol. 90, pp.3–12.
- Clary, P.M., Abboud, F.M., Somers, V.K. and Dyken, M.E. (1995) 'Sympathetic neural mechanisms in obstructive sleep Apnea', *J. Clin. Invest.*, Vol. 96, pp.1897–1904.
- Duda, R.O., Hart, P.E. and Stork, D.G. (2001) *Pattern Classification*, Wiley-Interscience Publication, New York.
- Flexer, A., Gruber, G. and Dorfner, G. (2005) 'A reliable probabilistic sleep stager based on a single eeg signal', *Artif. Intell. Med.*, Vol. 33, No. 3, pp.199–207.
- Flexer, A., Sykacek, P., Rezek, I. and Dorfner, G. (2001) 'An automatic, continuous and probabilistic sleep stager based on a Hidden Markov model', *Appl. Artif. Intell.*, Vol. 16, No. 3, pp.199–207.
- Furlan, R., Porta, A., Costa, F., Tank, J., Baker, L., Schiavi, R., Robertson, D., Malliani, A. and Mosqueda-Garcia, R. (2000) 'Oscillatory patterns in sympathetic neural discharge and cardiovascular variables during orthostatic stimulus', *Circulation*, Vol. 101, No. 8, pp.886–892.
- Kohavi, R. and John, G.H. (1997) 'Wrappers for feature subset selection', *Artificial Intelligence*, Vol. 97, Nos. 1–2, pp.273–324.
- Lewicke, A., Sazonov, E., Corwin, M.J., Neuman, M., Schuckers, S. and CHIME Grp. (2008) 'Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models', *IEEE Trans. Biomed. Eng.*, Vol. 55, No. 1, pp.108–118.

- Mainardi, L.T., Bianchi, A.M., Furlan, R., Piazza, S., Barbieri, R., di Virgilio, V., Malliani, A. and Cerutti, S. (1997) 'Multivariate time-variant identification of cardiovascular variability signals: a beat-to-beat spectral parameter estimation in vasovagal syncope', *IEEE Trans. Biomed. Eng.*, Vol. 44, No. 10, pp.978–989.
- Penzel, T., Kantelhardt, J.W., Lo, C.C., Voigt, K. and Vogelmeier, C. (2003) 'Dynamics of heart rate and sleep stages in normals and patients with sleep Apnea', *Neuropsychopharmacology*, Vol. 28, pp.48–53.
- Penzel, T., McNames, J., Murray, A., de Chazal, P., Moody, G. and Raymond, B. (2002) 'Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings', *Med. Biol. Eng. Comput.*, Vol. 40, No. 4, pp.402–407.
- Porée, F., Kachenoura, A., Gauvrit, H., Morvan, C., Carrault, G. and Senhadji, L. (2006) 'Blind source separation for ambulatory sleep recording', *IEEE Trans. Inf. Technol. Biomed.*, Vol. 10, No. 2, pp.293–301.
- Principe, J.C., Gala, S.K. and Chang, T.G. (1989) 'Sleep staging automaton based on the theory of evidence', *IEEE Trans. Biomed. Eng.*, Vol. 36, No. 5, pp.503–509.
- Rabiner, L.R. (1989) 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, Vol. 77, pp.257–286.
- Rechtschaffen, A. and Kales, A.E. (1968) *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, US Government Printing Office, NIH Publication No. 204, Washington DC.
- Redmond, S.J., de Chazal, P., O'Brien, C., Ryan, S., McNicholas, T.W. and Heneghan, C. (2007) 'Sleep staging using cardiorespiratory signals', *Somnologie*, Vol. 11, pp.245–256.
- Report of American Academy of sleep Medicine Task Force (1999) 'Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research', *Sleep*, Vol. 22, No. 5, pp.667–689.
- Rhoades, R.A. and Tanner, G.A. (2003) *Medical Physiology*, 2nd ed., Lippincott Williams and Wilkins.
- Scholz, U.J., Cerutti, S., Kubicki, S. and Bianchi, A.M. (1997) 'Vegetative background of sleep: spectral analysis of the heart rate variability', *Physiol. Behav.*, Vol. 62, No. 5, pp.1037–1043.
- Sforza, E., Jouny, C. and Ibanez, V. (2000) 'Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response', *Clin. Neurophysiol.*, Vol. 111, No. 9, pp.1611–1619.
- Sforza, E., Pichot, V., Barthelemy, J.C., Haba-Rubio, J. and Roche, F. (2005) 'Cardiovascular variability during periodic leg movements: a spectral analysis approach', *Clin. Neurophysiol.*, Vol. 116, No. 5, pp.1096–1104.
- Somers, V.K., Dyken, M.E., Mark, A.L. and Abboud, F.M. (1993) 'Sympathetic nerve activity during sleep in normal subjects', *N. Engl. J. Med.*, Vol. 328, pp.303–307.
- Somers, V.K., Dyken, M.V. and Skinner, L.J. (1993) 'Autonomic and hemodynamic responses and interactions during the Mueller maneuver in humans', *J. Auton. Nerv. Syst.*, Vol. 44, pp.253–259.
- Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) 'Heart rate variability: standards of measurement, physiological interpretation, and clinical use', *Circulation*, Vol. 93, pp.1043–1065.
- Todorova, A., Hofmann, H.C. and Dimpfel, W. (1997) 'A new frequency based automatic sleep analysis-description of the healthy sleep', *Eur. J. Med. Res.*, Vol. 2, pp.185–197.
- Watanabe, T. and Watanabe, K. (2004) 'Noncontact for sleep stage', *IEEE Trans. Biomed. Eng.*, Vol. 51, No. 10, pp.1735–1748.